Multi-Associative Memory in fLIF Cell Assemblies

Christian R. Huyck (c.huyck@mdx.ac.uk)

Kailash Nadh (k.nadh@mdx.ac.uk)

School of Engineering and Information Sciences
Middlesex University
London, UK

Abstract

The fundamental mammalian behaviours of perception, recognition, recollection, and all other psychological phenomena are intrinsically related to the basic cognitive tasks of memorisation and association. Based on Hebb's Cell Assembly (CA) theory, it is believed that concepts are encoded as neuronal CAs in mammalian cortical areas. This paper describes a series of simulations that demonstrate various associative memory tasks using CAs based on biologically plausible fatiguing, Leaky, Integrate and Fire neurons. The simulations show the ability of CAs to form, retain and recollect basic concepts and multiple and sequential associations.

Keywords: Cell Assemblies; Multi-associative memory; fLIF neurons

Introduction

Associative memory is a fundamental cognitive process. The concepts in memory and associations between them, are learned. These concepts and associations are critical to cognitive processing.

Like all cognitive processes, associative memory must have a neural basis, but neural models of associative memory are rare and surprisingly incomplete. Cell Assemblies (CAs) can account for many cognitive phenomena, including associative memory. Concepts can be stored as CAs (see Section CAs and auto-associative memory), and associations can be stored in connections between CAs.

Associative memory has a wide range of properties. Concepts can be connected in one to one, one to many, and many to many relationships. Associations can be context sensitive. In this paper, simulated CAs are used to explore these properties performing different tasks including a simple spatial cognitive mapping task. Cognitively, a good associative memory model should be capable of priming, differential associations, timing, gradual learning and change, encoding instances, and many such processes. The model simulations do not account for these phenomena, but this is the beginning of an exploration of a model that will (see Section Discussion and conclusion).

Background

Human associative memory is remarkable. Throughout life, new concepts are learned and new associations formed. Any given concept is associated with many other concepts, and retrieval of an associated concept can be based on a combination of the base concept and the context. There are a wide range of possible associations between concepts that have varying strengths, as is shown by, for example, priming studies. Memory retrieval and formation of associations are rapid processes.

Simulated neural models of associative memory are not currently capable of many of the tasks described in the prior paragraph. Closely related connectionist models have however been used to perform some of them.

CAs and auto-associative memory

Hebb (1949) hypothesised that the CA is the neural basis of concepts, and the CA is central to most neural models of memory. The theory proposes that objects, ideas, stimuli and even abstract concepts are represented in the brain by simultaneous activation of large groups of neurons with high mutual synaptic strengths (Wennekers & Palm, 2000). If an external stimulus excites a sufficient number of neurons of an existing CA, it can result in the spreading of activation within the CA, in turn igniting it due to recurrent activity and high mutual synaptic strength. The CA can remain active even after the stimulus is removed. This reverberating behaviour accounts for short term memory.

CAs are learned using the Hebbian learning rule, whereby modifications in the synaptic transmission efficacy are driven by the correlations in the firing activity of pre-synaptic and post-synaptic neurons (Gerstner & Kistler, 2002). When external stimuli are presented to a network, synaptic strength between neurons are adjusted so as to gain more strength if they undergo repeated and persistent activation or firing, gradually assembling them into a group, a CA. This formation of CAs accounts for long term memory. Thus, the CA hypothesis provides a structural and functional account for such cortical processes.

While still unproven, there is significant evidence and wide spread agreement that CAs are the neural basis of concepts. This includes a range of neural recording mechanisms (Abeles, Bergman, Margalit, & Vaddia, 1993; Bevan & Wilson, 1999; Pulvermuller, 1999).

The CA is a form of auto-associative memory. In auto-associative memories, an initial state is allowed to settle into a stored memory, allowing subsequent noisy input to retrieve a stored pattern. The Hopfield Model illustrates this property (Hopfield, 1984). A network of units that are well connected with bidirectional weighted connections is used to store a set of binary patterns (typically using a Hebbian calculation). When an initial set of neurons is switched on, in the discrete

version of the system, activation spreads through the system based on the weighted connections. In most cases the system will settle into a stable state with no neurons switching between on and off. If the input pattern is close to a stored pattern, it will settle into that pattern's state, thus functioning as a content-addressable memory. Neurons may also belong to multiple CAs. Hopfield patterns that share on-bits are models of CAs that share neurons.

While CAs are critical for the model of multi-associative memory described in this paper, they are not the solution. The question is how different CAs are associated with each other.

Multi-associative memory

Auto-associative memory is not typically what is meant by associative memory. Instead, associative memory is generally a shortened form (usually implicitly) of multi-associative memory; this has also been called hetero-associative memory. Psychologically, memories are not stored as individual concepts, but large collections of associated concepts that have many to many connections (Anderson & Bower, 1980). Each memory (CA) is associated with many other memories (CAs).

CAs and multi-associative memory

Even though CAs account for memory formation, their precise neural dynamics are far from perfectly understood. As explained in the Section CAs and auto-associative memory, neurons in a network may belong to different CAs, and if they are repeatedly co-activated by different versions of the same stimulus, they tend to become associated (Hebb, 1949). This is based on the notion that events that occur together repeatedly should somehow belong together. Wennekers and Palm (2000) explained that every time these events occur in conjunction, they drive certain subgroups of neurons, their correlated firing should be learned, and, by that, respective groups should become associatively connected.

Repeated co-activation of neurons can lead to the formation of CAs. Similarly, repeated co-activation of multiple CAs result in the formation of multiple and sequential associations, and sometimes new CAs. When an external stimulus activates a CA, it might lead to the activation of neurons that ignites a different CA that is not directly stimulated. This forms the rudimentary, neural level explanation of associative memory. Humans constantly retrieve and form associations with whatever sensory input they receive for the purpose of perception, understanding and reasoning.

Multi-associative memory models

Many multi-associative memory models have been proposed. A select few models are reviewed below.

Non-Holographic Associative Memory is an early multi-associative memory model (Willshaw, Buneman, & Longuet-Higgins, 1969). It is a well-connected network that can learn to map input bit patterns to output bit patterns using a Hebbian learning mechanism. In CA terms, input CAs are connected to output CAs via learned one way associations. This is a one step model.

The Linear Associator is a simple feed-forward, unidirectional neuronal model where patterns are stored associatively and retrieved by presenting linked patterns. A Linear Associator network can produce a near accurate associative output despite a partial or erroneous input pattern, exhibiting fault tolerance (Kohonen, 1977). The Linear Associator and many other models encode memories in well connected systems. The brain is not well connected, but it is often argued that it is broken into compartments that are well connected (Amit, 1989).

The **Multi Modular Associative Memory** (Levy & Horn, 1999) took advantage of well connected modules and analysed the storage capacity of a system with items stored in multiple modules. It showed that such a multi modular network is resilient to corrupted input, based on their observation that natural associated memories remain resilient to a great extent in humans who suffer from focal damage. They concluded that multi modular networks are necessary for meaningful implementation of associative neural networks. This is supported by evidence that shows that the memory for a given word is stored in multiple areas of the brain (Pulvermuller, 1999).

The **Valiant model** (Valiant, 2005) is a graph theoretical model of memorisation and association based on four quantitative parameters associated with the cortex: the number of neurons per concept; number of synapses per neuron; synaptic strengths; and number of neurons in total. It is assumed that neurons are randomly connected. The learning algorithm provided is biologically implausible, but the model shows that random graphs allow a method of assigning new memory items and associative relationships between the items.

The **Jets and Sharks** simulation (McClelland, 1981) uses the interactive activation model (Rumelhart & McClelland, 1982) to simulate associative memory. In the model, each concept is represented by a node, and connections are made between nodes to show how closely related these are. The system is not well connected. Activation spreads between the nodes via the weighted connections. The information to be encoded concerns two hypothetical groups (*Jets* and *Sharks*), group members, and some of their demographic characteristics. This shows that the system can act as a content-addressable memory system. So, the features of an individual group member can be activated as input, and the individual's representation will quickly become activated by the spread of activation. Additionally, prototypical effects can be derived (Rosch & Mervis, 1975). So, if the *Shark* concept is stimulated, activation will spread and eventually, the prototypical shark will become more active than other individuals. The individual that shares most features with other *Sharks* is the prototypical member.

This has been a brief review of multi-associative memory models. It has been known for 40 years that simulated neural systems can encode multi-associative memories. It has become apparent that well connected systems are not a good model of the brain. This has been addressed by partition-

ing the system into modules, and by using sparsely connected random graphs. These models however do not account for a range of associative memory characteristics that the human memory system exhibits, for instance, context effects.

The simulator

This section briefly describes a computational model that simulates CAs using fLIF neurons. Akin to all existing models, it is a simplification of the mammalian neural architecture, but has proven successful in modelling many cognitive phenomena.

The fLIF neural network

The fLIF neuron model (Huyck, 2007) encompasses many properties of the biological neuron. The CAs used in the experiments described in this paper emerge from fLIF neural networks. The model is an extension of the LIF (Leaky Integrate and Fire) model (Maas & Bishop, 2001; Gerstner, 2002). fLIF neurons collect activation from pre-synaptic neurons and fire on surpassing a threshold, that is, they integrate and fire. On firing, a neuron loses its activation level, otherwise the activation leaks gradually, resembling the behaviour of a biological neuron.

The activation A of a neuron i at time t is:

$$A_{i_t} = \frac{A_{i_{t-1}}}{\delta} + \sum_{j \in V_i} w_{ij} s_j \quad (1)$$

The current total activation is the remnant activation from the last time step divided by decay factor δ, plus incoming activation. This new activation is the sum of the active inputs s_j of all neurons $j \in V_i, V_i$ being the set of all neurons connected to i, weighted by the connection from neuron j to i. The neuron fires when the accumulated activation A exceeds a threshold θ. Firing is a binary event, and activation of w_{ij} is sent to all neurons j to which the firing neuron i has a connection. Fatiguing causes the threshold to be dynamic, $\theta_{t+1} = \theta_t + F_t$. F_t is positive (F_+) if the neuron fires at t and negative (F_-) if it does not.

The network architecture

Two of the three the simulations discussed in this paper partitions the network into subnetworks; the context simulation uses only one subnet. The subnets are made of fLIF neurons and the number of neurons vary between subnets. Intra-subnet synapses are based on biologically inspired distance biased connections. This topology makes it likely for a neuron to have excitatory connections to neighbouring neurons, and less likely to far away ones. The subnet is a rectangular array of neurons with distance organized toroidally. Inhibitory connections within a subnet and all inter-subnet connections are set up randomly. The connectivity rule for excitatory neurons is given by equation 2. There exists a connection between neuron i and j of a network only if $C_{ij} = 1$.

$$\begin{aligned} C_{ij} &= 1, \text{if } r < (1/(d * v)) \quad (2) \\ C_{ij} &= 0, \text{if not} \end{aligned}$$

where r is a random number between 0 and 1, d is the neuronal distance and v is the connection probability. This indicates that connections in a network are influenced by distance between neurons and the connection probability factor. Distance $d = 5$ throughout all the simulations, as it has been observed to work well. Inspired by the biological neural topology, long distance intra-network connections are also present, connected by long distance axons with many synapses (Churchland & Sejnowski, 1992).

In each of the simulations, networks are divided into multiple CAs in response to stimuli using unsupervised learning algorithms. The CAs are orthogonal and represent different concepts, and this is based on training. Neurons in different CAs do have excitatory connections to other CAs, based on the connection rule (Equation 2), but the learned weights are low because neurons in different CAs rarely co-fire. Once learned, when a CA is externally activated, it inhibits all inactive CAs in the same network via learned inhibitory connections. Similarly, simultaneous co-activation of CAs increases the connection strength between them, creating associations.

Learning in the network

CAs in a network are learned by a correlatory Hebbian learning rule (Huyck, 2004), whereby synaptic connection weights are modified based on the following equation:

$$\begin{aligned} \Delta^+ w_{ij} &= (1 - w_{ij}) * \lambda \quad (3) \\ \Delta^- w_{ij} &= w_{ij} * -\lambda \quad (4) \end{aligned}$$

w_{ij} is the synaptic weight from neuron i to j and λ is the learning rate. During each cycle, weights change based on the state of pre-synaptic and post-synaptic neurons. If both neurons fire, the weights increase as per the Hebbian rule (Equation 3). If only the pre-synaptic neuron fires, weights decrease as per the anti-Hebbian rule (Equation 4). These two rules act together, changing w_{ij}, gradually increasing the likelihood of j firing if i fires. Without reverberation, the weight would reflect the likelihood that neuron j fires when neuron i fires.

The network parameters used in the simulations are presented in the table below:

Table 1: Network parameters

Parameter	Symbol	Value
Learning rate	λ	.10
Activation threshold	θ	4.5
Fatigue	$F_+ = F_-$	.80
Decay factor	δ	1.2
Neuronal distance	d	5

Simulations

This section describes three sets of simulations. These simulations demonstrate that the model is capable of supporting complex associations.

Jets and Sharks

This is a CA based implementation of a modified version of the classic Jets and Sharks model that uses five members in each of the hypothetical (*Jets* or *Sharks*) groups. Each of the members and their attributes are encoded as CAs in different subnets. A unique *Person* CA represents each member and their attributes, namely (*Name*, *Age*, *Education*, *Marital status*, and *Occupation*). There is a one-to-one relationship between each *Person* and their *Name* CA. A subset is illustrated in Figure 1.

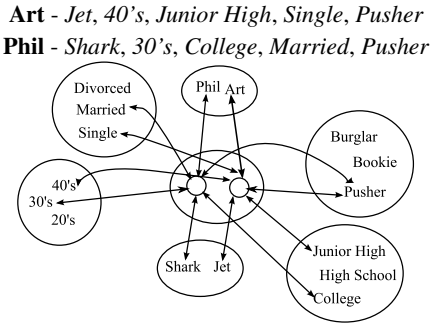

Figure 1: Network setup of Jets and Sharks subset. Circles refer to subnets, names to orthogonal CAs, and arrows to connections.

Each CA is mutually exclusive and is made up of 200 neurons. Inter-subnet connections are random, initially low-weight excitatory connections. The CAs and their associations are learned by external stimulation and co-activation of each *Person* CA and their attributes simultaneously for 200 cycles, in succession.

Multiple memory retrieval tests were conducted. For instance, when the *Name* CA of **Art** is externally stimulated, it propagates activation to *Art*'s *Person* CA. The particular *Person* CA, having learned excitatory connections to different attributes, causes activation to further propagate throughout the network, gradually activating all corresponding attributes of **Art**. On 15 runs, the correct results were retrieved for each of the 10 people.

Similarly, when the attribute *Shark* is externally activated, it propagates activation to all *Person* CAs having that attribute, and the immediate effect is that all *Shark* members come on. Gradually, activations stabilise and one *Person* CA is found to have more activation than others, emerging as the prototypical *Shark*. The network was tested 15 times to obtain the prototypical *Jet* and **Art** emerged to be so, throughout. The same was done for *Sharks* and **Nick** emerged to be prototypical 9 times and **Ned**, 6 times. This is because both members share the most features with other members of the group, and hence emerge to be prototypical members.

Context sensitive association

Most associative memory models, focusing solely on associations, usually neglect to acknowledge the inherent *types* of associations that exist. A concept may be associated to many others, but the types of associations may vary from concept to concept. The association of *cat* to *mammal* may not be the same as *fur* to *mammal*. As an initiative towards simulating differentiable associations and eventually implicit labelled associations, a model capable of differentiating associations based on contexts was developed.

The parameters in the simulation are those from Table 1, except the fatigue parameters have been modified. $F_+ = F_- = 0.4$.

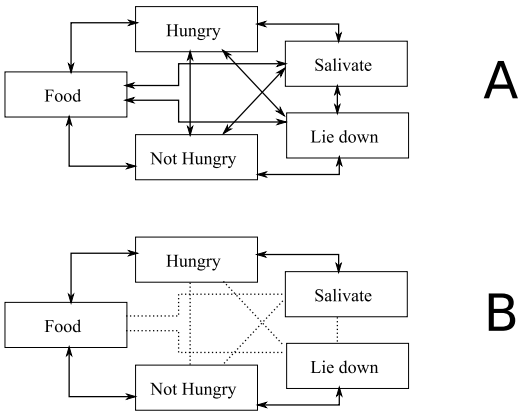

Figure 2: Initial and Learned state of CAs

Figure 2 shows the network setup, the physical connection before (A) and after (B) the CAs and their associations are learned. A single network holds all the 5 orthogonal CAs, namely *Hungry, Not Hungry* (states), *Salivate, Lie down* (actions) and *Food* (object). Since all the CAs are in the same network, they have excitatory and inhibitory connections with each other. Initially, patterns corresponding to each of the CAs are presented for 300 cycles so that they are learned independently. When a CA is active, it inhibits all other CAs in the network via learned inhibitory connections. When one CA is active and another is inactive, inter-CA connection weights are decreased. The associations between CAs are learned by co-activation for 300 cycles each, that is, by activating three CAs (object, state, action) simultaneously at a time, in the following manner:

$$\textit{Food} + \textit{Hungry} \Rightarrow \textit{Salivate}$$
$$\textit{Food} + \textit{Not Hungry} \Rightarrow \textit{Lie down}$$

This mimics the behaviour of a hypothetical *dog* that salivates when food is presented when hungry, and lies down ignoring food when not hungry. After the associations are learned, context sensitive behaviour is tested in the following manner: when *Food* and *Hungry* are externally stimulated, *Salivate* activates, suppressing *Lie down*. Similarly, when *Food* and *Not Hungry* are externally stimulated, *Lie Down* activates, suppressing *Salivate*.

The tests were repeated on 100 different network configurations, and action CAs (*Salivate, Lie down*) activated correctly 83 times with an average of 84.6 neurons firing.

Cognitive spatial mapping using sequential memory

Cognitive spatial mapping is a psychological process by which an individual acquires, stores, recalls and decodes information about the relative locations and attributes of a spatial environment for the purpose of spatial navigation (Downs & Stea, 1973).

A simplified version of this complex process was implemented, where a virtual agent navigates a 3D virtual world by recognising, memorising, associating and recollecting rudimentary landmarks. The parameters in the simulation are those from Table 1, except the fatigue parameters have been modified. $F_+ = F_- = 0.1$.

Figure 3 shows the top view of the virtual world, its 4 rooms, 4 coloured doors, and the exploration path the agent takes. The agent's path is fixed and it lacks the ability to turn back and only moves forward.

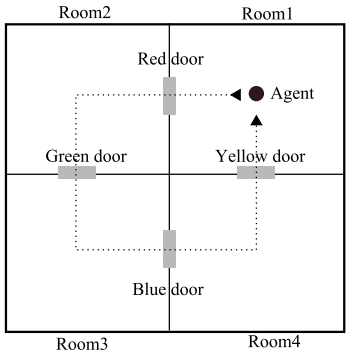

Figure 3: Top view of the virtual world

In the learning mode, the agent explores the world, learning rooms, doors and *Room-Door-Room* sequences in the process. A simple vision system detects doors and door colours, and triggers learning actions, helping the agent navigate.

Figure 4 illustrates the network setup of the spatial mapping module, excluding other subnets of the agent. The greyed areas show a sample sequence, where the agent has learned the association *Room1*$\longleftrightarrow$*Red-Door*$\longleftrightarrow$*Room2* by co-activation. **RoomNet1** and **RoomNet2** store instances of the rooms the agent visits. **DoorNet** stores the doors encountered and **SequenceNet**, encodes the sequences of visits. **ColourNet** has CAs that represent colours recognised by the agent, and **GoalNet** encodes the target door, which the agent searches for while in the test mode. The arrows in the figure show physical inter-subnet connections (random, low-weight excitatory connections). The CAs representing rooms, doors and sequences in corresponding subnets are made up of 200 neurons each, and are learned as the agent explores.

In the learning mode, when the agent encounters a door, a 5-step learning process is triggered: 1) The agent encodes its present location as a CA in both the **RoomNets**; 2) It learns the door, forming a CA in the **DoorNet**; 3) The agent associates the colour of the door in the **ColourNet** with the newly formed door CA; 4) The agent moves to the next room and

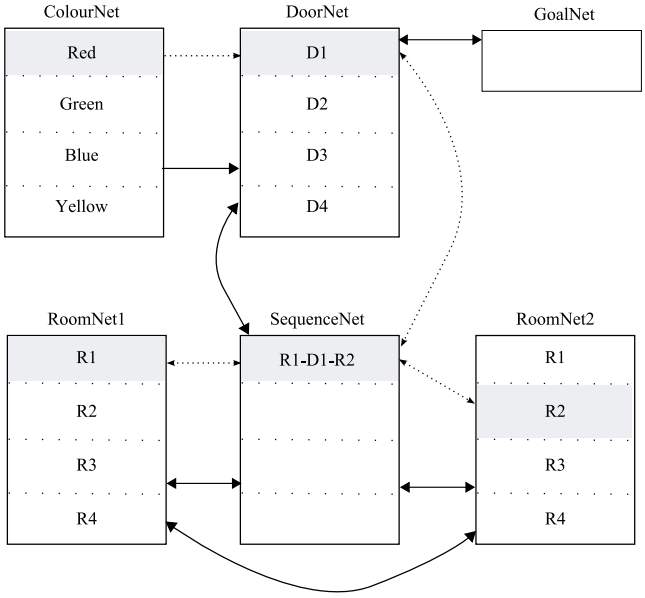

Figure 4: Cognitive spatial mapping network setup

learns the room (as in Step 1); 5) It then encodes the passage it just made as a CA in the **SequenceNet**, as in *PreviousRoom-ConnectingDoor-PresentRoom*. This process is repeated until the agent is back at its starting position. Each of the CAs are learned by stimulation lasting 300 cycles, triggered by the visual cues the agent reviews. Associations are learned by co-activation, as described in the previous simulations. For instance, passages are learned by simultaneously activating the corresponding sequence CA in the **SequanceNet**, preentrance room CA in **RoomNet1**, the connecting door CA in **DoorNet** and the present room CA in **RoomNet2**, for 300 cycles. In the test mode, the agent is instructed to go to a randomly chosen room of the 4 rooms. This is done by externally stimulating the target room CA in *RoomNet2*. This causes the corresponding sequence CA to activate, which in turn activates the associated room CA in **RoomNet1** and the connecting door CA in **DoorNet**. When the door CA becomes active, the goal CA is activated externally, leaving them to remain simultaneously active for 300 cycles, causing them to become associated. As a result of this association, the goal CA becomes active whenever the corresponding door CA activates. The active door CA that the agent has set as its goal is the door that leads to the target room. With the goal in memory, the agent moves forward, looking for the target landmark (door). When the target door appears in the agent's visual field, the corresponding door CA in **DoorNet** activates, immediately causing the goal CA to activate due to the previously learned association, indicating achievement of the goal. With this, the agent reaches the target room and the test ends.

The agent could have failed by stopping prematurely, or continuing beyond the target room. However, it correctly reached a randomly chosen target room all 30 times the test was repeated.

Discussion and conclusion

These simulations show that fLIF neural models are capable of learning and retrieving core memories, in the form of CAs, and associations between them. The Jets and Sharks simulations show that the model can handle one to one, one to many, and many to many relations. The second set of simulations show that the model is capable of handling context sensitive associations, and the third set shows that it is capable of a basic form of cognitive mapping, using multi-associative sequential memories. This is the first neural model, that the authors are aware of, that simulates all the above mentioned processes.

While these are useful capabilities, the model does not exhibit the wide range of behaviours that human associative memory does. Human memories have varying strengths, and so do the associations. Instances of types (tokens) can be learned. Types, associations and tokens are all forgettable. All of these behaviours occur in measurable times. It is planned that future work will include all of these behaviours. In the simulations described in this paper CAs were orthogonal, that is, neurons were in only one CA. Associations were maintained solely by synapses between neurons in the associated CAs. Another type of association is possible, where CAs are associated by sharing common neurons. Subcategorisation associations have been stored using shared neurons in CAs (Huyck, 2007) . For example, the concept *Cat* shares neurons with the concept *Mammal* because of the association that a *Cat isA Mammal*. It is likely that such overlapping CAs are important for a good neural implementation of multi-associative memory.

Other properties may also be necessary to achieve the full range of associative memory behaviours. For instance, global inhibitory mechanisms might be needed to manage spreading of activation and prevent all neurons firing simultaneously (simulated epilepsy). None the less, the current simulations show that simulated neural systems can perform a wide range of associative memory tasks. The authors leave the reader with this question: what tasks does an associative memory perform?

References

Abeles, M., Bergman, H., Margalit, E., & Vaddia, E. (1993). Spatiotemporal firing patterns in the frontal cortex of behaving monkeys. *Journal of Neurophysiology*, *70(4)*, 1629–1638.

Amit, D. (1989). *Modelling brain function: The world of attractor neural networks*. Cambridge University Press.

Anderson, J. R., & Bower, G. H. (1980). *Human associative memory: A brief edition*. USA: Lawrence Erlbaum Associates.

Bevan, M., & Wilson, C. (1999). Mechanisms underlying spontaneous oscillation and rhythmic firing in rat subthalamic neurons. *Neuroscience*, *19*, 7617–7628.

Churchland, P., & Sejnowski, T. (1992). *The computational brain*. USA: MIT Press.

Downs, R. M., & Stea, D. (1973). Cognitive maps and spatial behaviour. process and products. In *Image and environment: Cognitive mapping and spatial behaviour* (pp. 8–26). Chicago: Aldine.

Gerstner, W. (2002). Integrate-and-fire neurons and networks. In *The handbook of brain theory and neural networks* (pp. 577–581). The MIT Press.

Gerstner, W., & Kistler, W. K. (2002). Mathematical formulations of hebbian learning. *Biological Cybernetics*, *87*(5-6), 404–415.

Hebb, D. O. (1949). *The organization of behavior*. New York: Wiley.

Hopfield, J. J. (1984). Neurons with graded response have collective computation properties like those of two-state neurons. In *Proceedings of the national academy of sciences* (Vol. 81, pp. 3088–3092).

Huyck, C. (2004). Overlapping cell assemblies from correlators. *Neurocomputing*, *56*, 435-9.

Huyck, C. (2007). Hierarchical cell assemblies. In *Connection science* (Vol. 19:1, pp. 1–24).

Kohonen, T. (1977). *Associative memory - a system theoretical approach*. Berlin: Springer-Verlag.

Levy, N., & Horn, D. (1999). Associative memory in a multimodular network. *Neural Computation*, *11*, 1717–1737.

Maas, W., & Bishop, C. M. (2001). *Pulsed neural networks*. MIT Press.

McClelland, J. L. (1981). Retrieving general and specific information from stored knowledge of specifics. In *Proceedings of the third annual conference of the cognitive science society* (pp. 170–2).

Pulvermuller, F. (1999). Words in the brain's language. *Behavioral and Brain Sciences*, *22*, 253–336.

Rosch, E., & Mervis, C. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, *7*, 573–605.

Rumelhart, D., & McClelland, J. (1982). An interactive activation model of context effects in letter perception: Part 2. the contextual enhancement and some tests and extensions of the model. *Psychological Review*, *89:1*, 60–94.

Valiant, L. G. (2005). Memorization and association on a realistic neural model. *Neural Computation*, *17*, 527–5557.

Wennekers, T., & Palm, G. (2000). Cell assemblies, associative memory and temporal structure in brain signals. In *Conceptual advances in brain research* (Vol. 2).

Willshaw, D., Buneman, O., & Longuet-Higgins, H. (1969). Non-holographic associative memory. *Nature*, *222*, 960–962.